Automatic Removal of Poor Quality Images from Digital Image Sequences

Gregory J. Power[a] and Mohammad A. Karim[b]

[a]Air Force Research Laboratory, AFRL/SNAT/Target Recognition Branch, 2010 5th Street, Wright-Patterson AFB, Ohio 45433-7001

[b]University of Dayton, Department of Electrical and Computer Engineering, 300 College Park, Dayton, Ohio 45469-0226

ABSTRACT

This paper introduces a spatio-temporal technique for selecting or filtering out lower quality digital image frames. The technique is demonstrated on Electro-Optical/Infrared image sequences which suggests it is a candidate for exploiting reconnaissance (recce) imagery or can be part of a recce subsystem. For human vision exploitation, a few poor quality image frames out of hundreds in a digital image sequence may be only a minor irritation when the sequence runs at the typical 30 frames per second. Of course, if that human needs to examine each frame, a system that automatically removes or enhances lower quality image frames could be beneficial. For machine vision subsystems, a few poor quality image frames could cause lower probability of recognition. The filtering technique introduced in this paper can improve input into machine vision algorithms. Another application for this technique is digital transmission to filter out unwanted images prior to transmission or to selectively enhance the poor quality frames. A major portion of current research into quality in digital image sequences focuses on transmission systems where an input high quality image sequence can be compared to the lower quality image sequence received at the output of the transmission system. However, this paper shows a technique for judging the quality of the input image frames prior to transmission, without a transmission system or without any knowledge of the higher quality image input. The impact of digital image artifacts on the spatio-temporal quality are shown. The quality variations in the individual frames of the input image sequence are charted to show which frames are of lower quality and thus need filtering.

Keywords: Target Recognition, Image Quality, Spatio-Temporal Analysis, Digital Transmission, Image Sequence, Recce subsystem, Reconnaissance, Velocital

1. INTRODUCTION

A technique is demonstrated that filters out poor quality images from a digital image sequence based on velocital information changes. More specifically, the digital image sequences are Electro-Optical/Infrared images that suffer from various sudden noise effects. The amount of noise that will be tolerated is selected by adjusting a threshold.

Conceptually, the technique is a candidate for exploiting reconnaissance (recce) imagery. At least, three possible applications for ground-based or airborne subsystems can be conceived: one for the human analyst, one for airborne transmission, and one for machine vision. On the ground, a subsystem that assists in automatically filtering out or enhancing poor quality image frames can reduce workload for the human analyst who needs to examine each frame.

In the air, a subsystem that removes or enhances lower quality image frames before transmission can reduce the bandwidth needed and length of transmissions. In some digital transmission systems, a frame difference and quantization is used to reduce the number of bits needed to transmit. Large erratic noise can increase the number of bits needed to transmit if an adaptive quantization scheme is used. As an example, consider the three sequential infrared image frames in Figure 1 which are 316 by 491 pixels and 8 bits per pixel. Using a frame difference, the maximum range of pixel values could be -255 to 255. A frame difference of the sequential infrared images is shown in Figure 2. In practice, with minor variations from frame to frame, it is hoped that a frame difference will

Other author information: (Send correspondence to G.J.Power)
G.J.Power: E-mail: powergj@sensors.wpafb.af.mil
M.A.Karim: E-mail: mkarim@engr.udayton.edu

predominantly produce a lower range of pixels and thus less quantization is needed. For the differences in Figure 2 the first and second frame results in a variance of 15.7 while the difference in the second and third frame results in a variance of 37. The larger variance requires more quantization levels to reproduce the image and thus the number of bits transmitted are increased for an image that may be of little value.

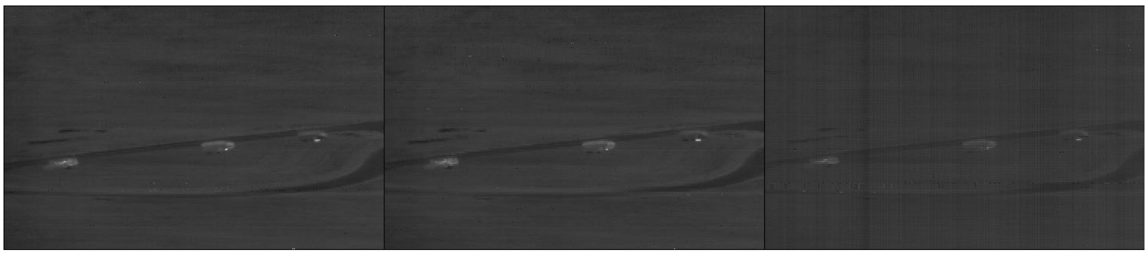

Figure 1. Three sequential frames from a digital infrared image sequence where degradation is evident in the third frame.

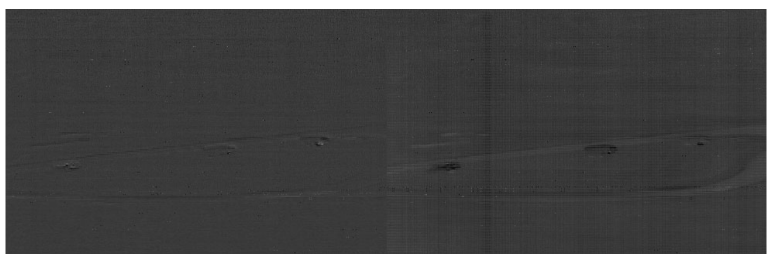

Figure 2. The left image is a the difference between the left two frames in Figure 1. The right image is the difference between the right two frames in Figure 1. The images have been scaled for visibility. The left image has an intensity variance of 15.7 while the right image has an intensity variance of 37.

Another conceptual use for filtering bad images is machine vision. For an object recognition system, the incorporation of a subsystem to remove or enhance images prior to entering the object recognition algorithms can enhance the probability of recognition. For a human observer who is observing frames passing at a rate of 30 frames per second, a few poor frames every now and then may be considered a minor irritation. After all, the human is able to mentally remove or intelligently integrate the few frames that hamper recognition. However, for machine vision, unless the object recognition algorithm incorporates some filter such as the one to be introduced by this paper, then a few bad frames can impact the overall probability of target recognition.[1] Consider a recognition model that is sequential such as shown in Figure 3. A high level statistical analysis of this model suggests there is a lower probability of recognition if there is lower input quality since

$$P(Recognition) = P(Match|Index, Feature, Segment, Detect, Quality).$$

So, the probability of recognition is the probability of getting a match given a certain amount of success in indexing, feature extraction, image segmentation, object detection and quality imagery. If the image quality is extremely poor, then all the steps in recognition can be hampered. This translates in layman terms to "garbage in, garbage out." If noise is erratic, then conceptually, an algorithm that eliminates or enhances the bad frames will reduce the "garbage in."

2. VELOCITAL INFORMATION FEATURE

By looking at a velocital information feature,[2] the identification of poor quality image frames can be recognized. The velocital information feature is defined as the standard deviation over the magnitude of the optical flow map on an image plane at a particular time. Letting, $v(x,y,t)$ represent image optical flow distribution over space x,y and time t with mean optical flow represented by μ_v and optical flow density function defined as

$$\begin{cases} f(x,y) & \text{if } 0 < x < x_{max} \text{ and } 0 < y < y_{max} \\ 0 & \text{otherwise} \end{cases}$$

then the velocital information feature is

$$\sigma_v(t) = \sqrt{\int_0^{y_{max}} \int_0^{x_{max}} (v(x,y,t) - \mu_v)^2 f(x,y) dx dy}. \tag{1}$$

$v(x,y,t)$ has been estimated by a model[3] where a point of image intensity, $I(x,y,t)$, is assumed not to change as it moves through space and time, i.e.,

$$\frac{dI}{dt} = 0.$$

Using this unchanging intensity model and partial differentiation with $\nabla I = [\frac{\partial I(x,y,t)}{\partial x} \frac{\partial I(x,y,t)}{\partial y}]^T$, the optical flow vector magnitude can be derived[3] as

$$v(x,y,t) = \frac{\frac{-\partial I(x,y,t)}{\partial t}}{\|\nabla I(x,y,t)\|}. \tag{2}$$

For a digital image sequence of M by N uniformly distributed discrete pixels, the velocital information feature of Equation 1 for a particular image frame in time, t_n, has been estimated discretely[2] as

$$VI_{stdev}[t_n] = \sqrt{\left[\frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} VI^2(x,y,t_n)\right] - VI_{mean}^2(t_n)},$$

in which

$$VI_{mean}(t_n) = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} VI(x,y,t_n)$$

with $VI(x,y,t_n)$ as the discrete version of Equation 2 representing the velocity magnitude at a particular pixel location. Power et al.[2] have shown a discrete computational implementation of $VI(x,y,t_n)$ using spatial and temporal information features.

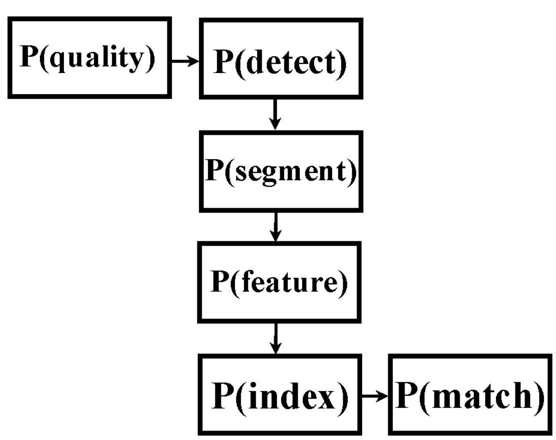

Figure 3. For a sequential model of object recognition, the probability of high recognition is dependent on input of sufficient quality.

3. EXPERIMENTAL SETUP

It has been shown that VIC[2] is slow to change based on camera functions such as zooming and panning but it is sensitive to image editing functions such as frame cuts. For an operational scenario where the camera is fixed and the data is received without editing at 30 frames per second, VIC is expected to be sensitive to erratic activity such as transient noise. Therefore for this test, infrared digital image sequences are used from a scenario where an airborne sensor is approaching (slowly zooming in to) a ground target. There are no edits and the infrared sequence suffers from an unknown noise content. The VIC metric was tested on various 3 second frame clips. All the data came from one sensor but different flights with different vehicle targets in different locations.

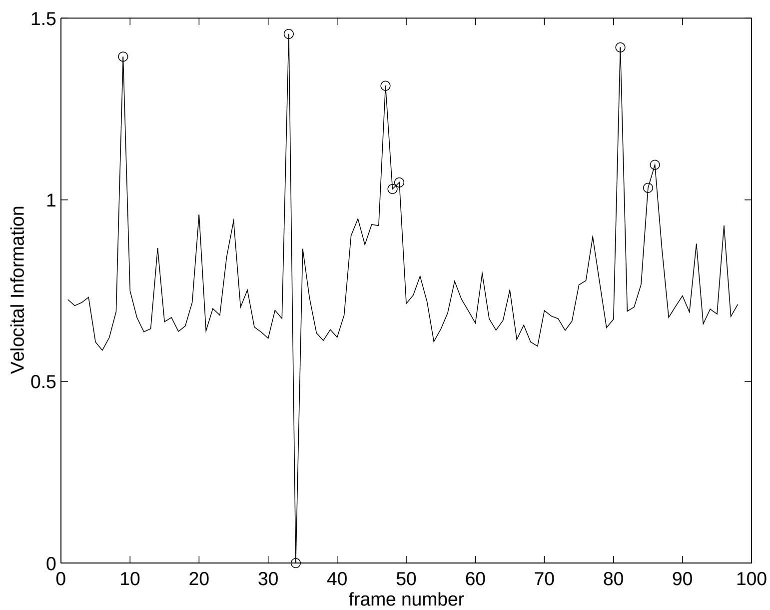

Figure 4. The velocital information content of 98 frames of an infrared image sequence. The poorest quality frames are circled including frames 9, 33, 34, 47, 48, 49, 81, 85 and 86.

4. RESULTS

A typical result of using VIC on infrared digital image sequences suffering from varying amounts of noise is shown in Figure 4 where 98 frames of sequential data are processed. The result is shown for a sequence that shows a temporal only problem caused by frame repetition (frames 33 and 34) and additional problems due to spatial and temporal noise degradation as shown in Figure 5. The poorest quality frames (9, 33, 34, 47, 48, 49, 81, 85, and 86) are circled. Since all the circled frames are obviously above all other frames, then a threshold can be chosen to remove the frames. Figure 6 shows a three dimensional spatio-temporal velocital plot of the same data with the same frames circled. It is obvious again that the frames are outside the expected range of variation for the image sequence. A plot of the spatial information in Figure 7 shows that the poorest frames could not have been chosen solely on spatial content. For this sequence, the velocital and spatial content has a .45 correlation coefficient. A plot of temporal information in Figure 8 shows more similarity to the velocital information with a .93 correlation coefficient indicating that the velocital content is tied closer to the temporal content for this particular sequence. The correlation coefficients for other 98-frame test sequences ranged from .25 to .7 for spatial-velocital correlation and .84 to .94 for temporal-velocital correlation. Figure 9 shows some of the frames selected as poorer quality along with the frame that preceded the poor quality frame.

In another test, a well-behaved sequence is used with good contrast. Still, as shown in Figure 10, VIC found one frame with noise content and three frame repeats. The frame repeats are good quality frames. The frame selected as poor quality is shown in Figure 11 along with the good quality frame just before it.

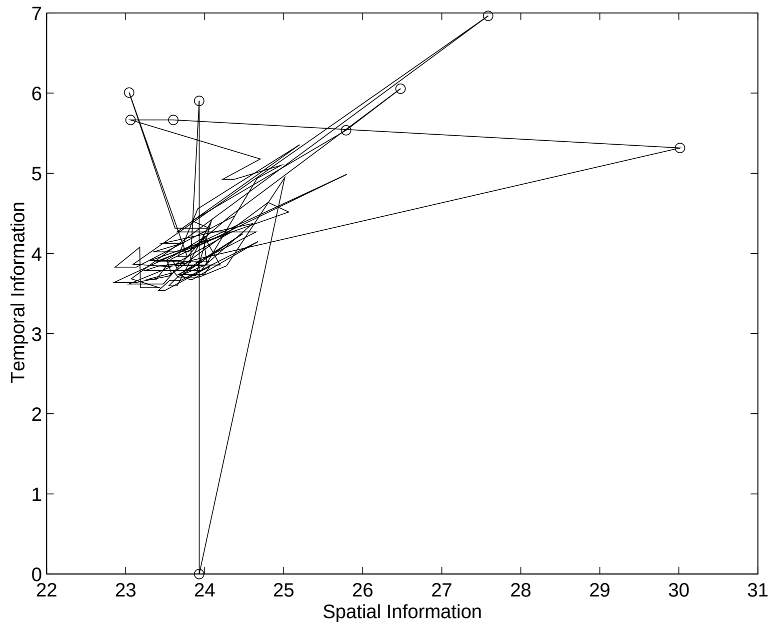

Figure 5. The spatio-temporal content of 98 frames of an infrared image sequence. The circled points are frames with poorest quality.

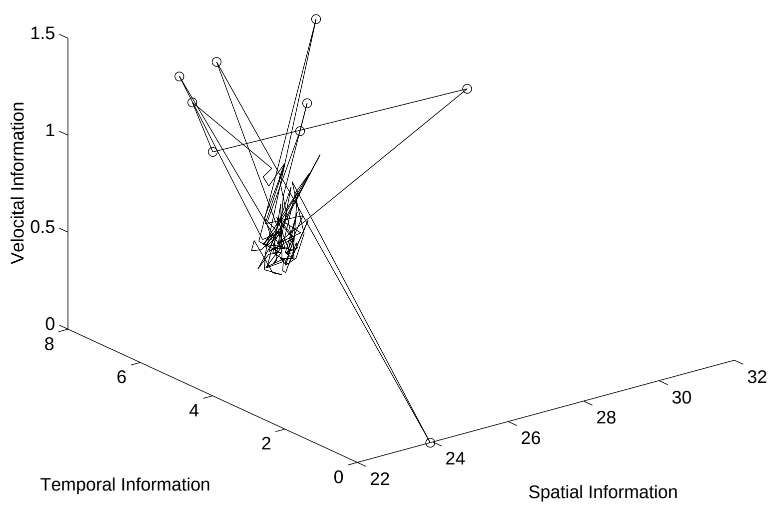

Figure 6. A plot of spatial, temporal, and velocital information for 98 frames of an infrared image sequence. The poorest quality frames as selected by VIC are circled and obviously leave the expected spatio-temporal region of variation.

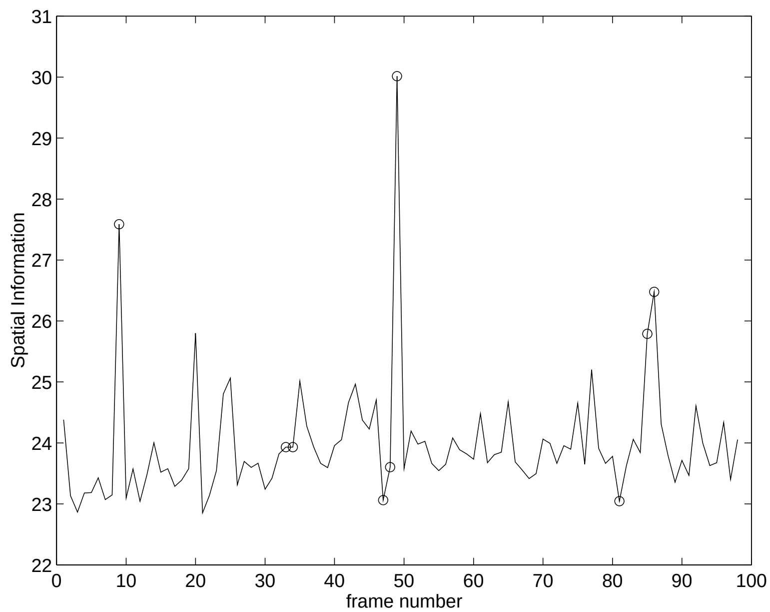

Figure 7. The spatial information content of 98 frames of an infrared image sequence. Spatial does not give sufficient information to select the poor quality frames.

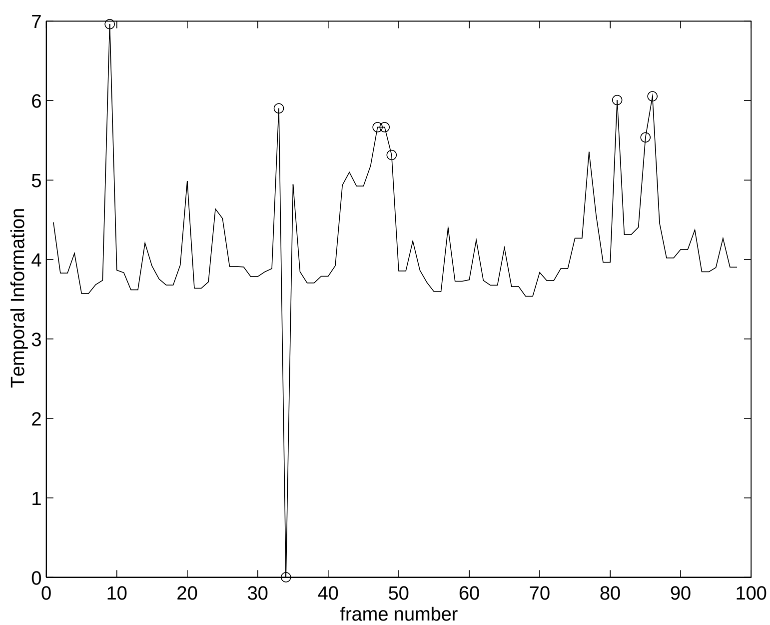

Figure 8. The temporal information content of 98 frames of an infrared image sequence. The poor quality frames selected based on velocital information are circled.

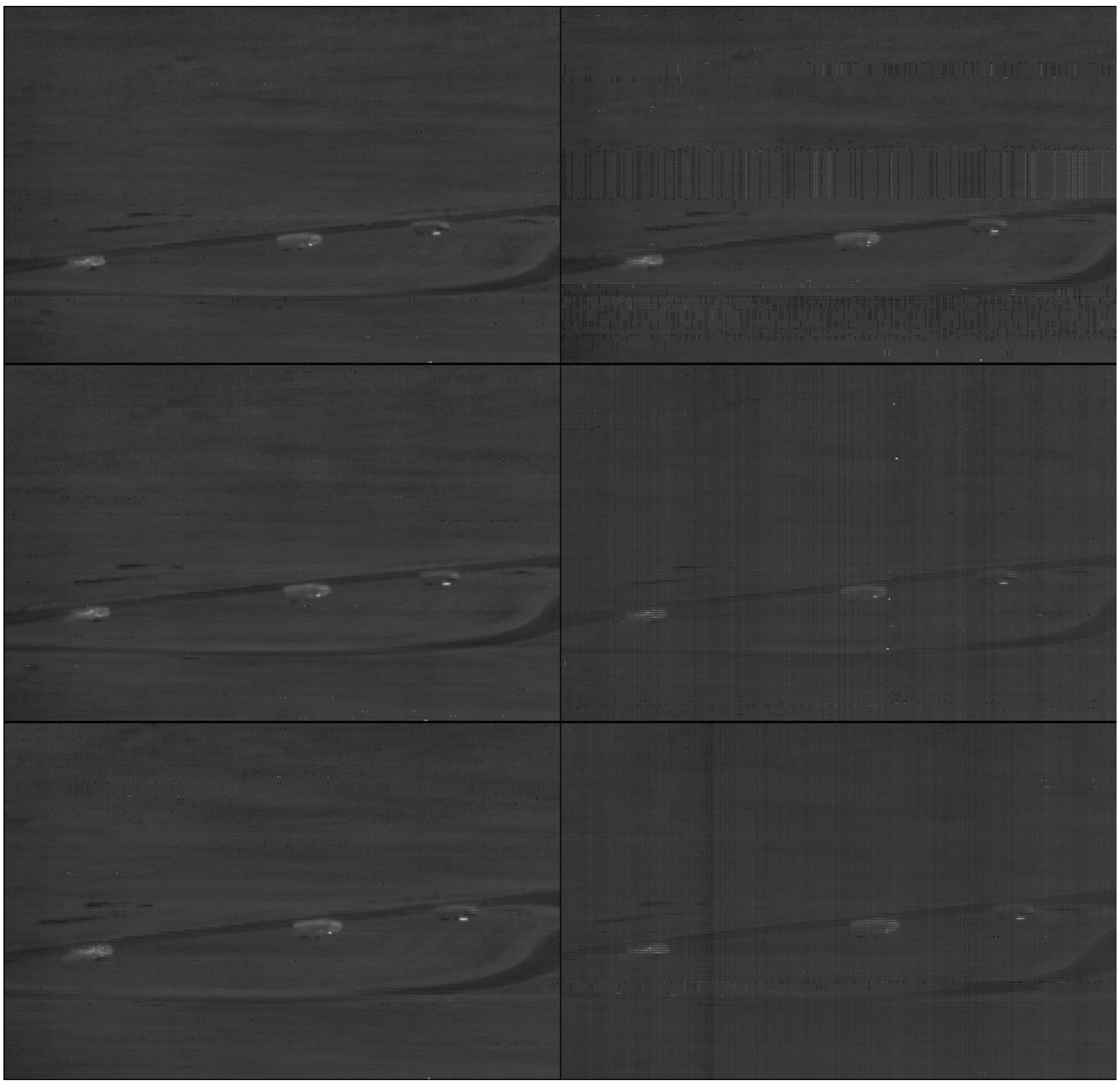

Figure 9. Using VIC, the right frames were selected as poorer quality. For comparison purposes, the left good frames are the frames one frame prior to the frame selected as poorer quality.

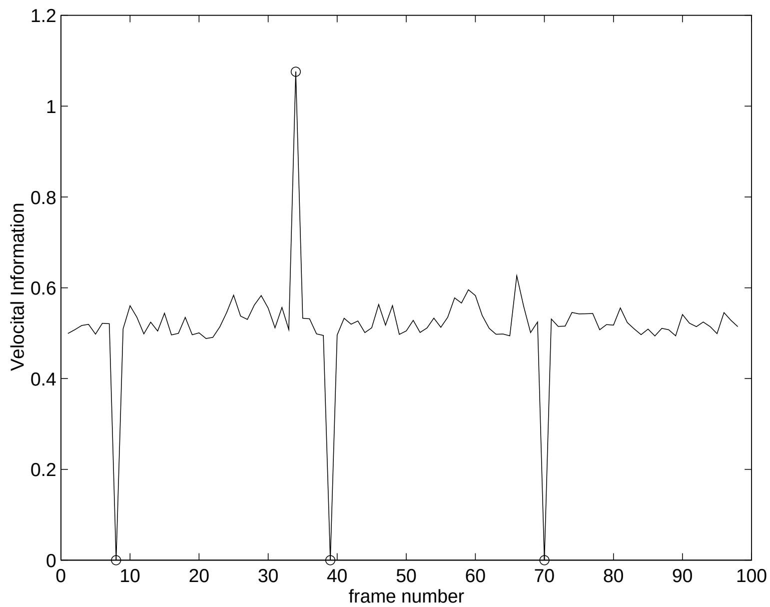

Figure 10. VIC for 98 frames of a good quality infrared image sequence shows less variation in VIC. Three frame repeats are found at frames 8, 39 and 70. A poor quality frame is flagged at frame 34.

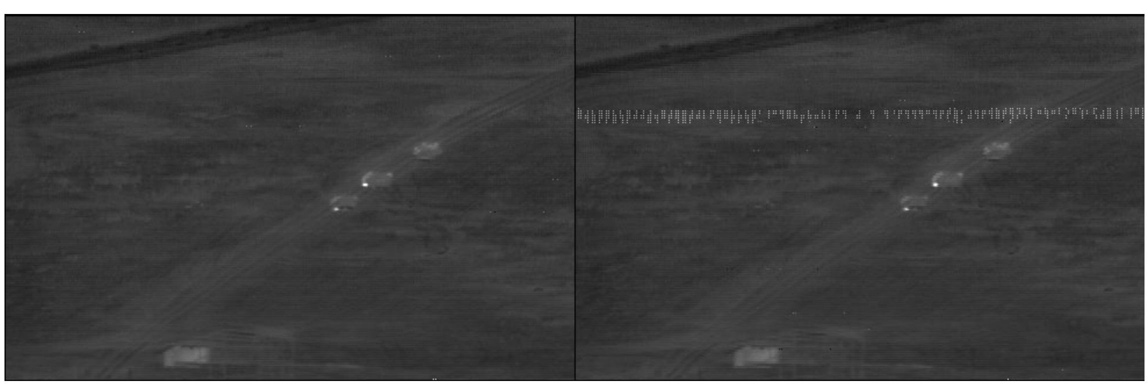

Figure 11. For the sequence charted in Figure 10, the left frame is the good quality frame, number 33, just before the poor quality frame , number 34, shown on the right.

5. CONCLUSION

A major portion of current research into quality in digital image sequences focuses on transmission systems where an input high quality image sequence can be compared to the lower quality image sequence received at the output of the transmission system. However, this paper shows a technique for judging the quality of the input image frames prior to transmission, without a transmission system or without any knowledge of the higher quality image input. Given digital infrared image sequences taken from an airborne platform, a velocital information metric based on spatio-temporal information is used to successfully select frames with various amounts of erratic noise and then automatically rejects the frames based on a predetermined threshold.

REFERENCES

1. G. J. Power and M. A. Karim, "Determining a confidence factor for automatic target recognition based on image sequence quality," in *Algorithms for Synthetic Aperture Radar Imagery V*, E. G. Zelnio *et al.*, eds., *Proc. SPIE* **3370**, April 1998.
2. G. J. Power, M. A. Karim, and F. Ahmed, "Charting image artifacts in digital image sequences using velocital information content," in *Applications of Digital Image Processing XXI*, A. G. Tescher *et al.*, eds., *Proc. SPIE* **3460**, July 1998.
3. A. M. Tekalp, *Digital Video Processing*, Prentice-Hall, Inc., Upper Saddle River, NJ, 1995.